

Measuring cross correlation function line-profile variations in radial-velocity measurements via a Skew Normal distribution

U. Simola

X. Dumusque

Jessi Cisewski-Kehe

March 15, 2018

Abstract

When working with radial velocities for detecting Extrasolar planets and when using data from stabilized spectrographs, the different moments of the cross-correlation function (CCF) are used to measure the radial velocity of the star but also the changes in the shape of the CCF. Those changes are due to stellar activity and therefore the best precision on these moments is required to de-correlate exoplanet signals from spurious RV signals originating from stellar activity. We propose here to measure those moments using a Skew Normal (SN) distribution, which compared to the Normal distribution generally used, provides an extra parameter to model the CCF natural asymmetry. We analyzed 5 stars with different activity levels and different signal-to-noise ratio levels. In each case, we compare the results obtained from the Normal fitting of the CCF and from the SN fitting. We also estimate rigorous errors for the different moments of the CCF using a bootstrap analysis. The correlations between the RVs and the CCF asymmetry or RVs and the CCF width is always stronger when using the parameters derived from the SN. Therefore all the moments of the CCF are more sensitive to activity, which is a huge gain to probe stellar activity. RVs derived from a SN distribution are more sensitive to activity, which is interesting when looking for rotational periods or characterizing better stellar activity. **[[Referee: "interesting" is a highly subjective and not very precise or quantitative expression. Interesting in what sense? I would term this a "sentence of no value", i.e. one that conveys not useful information to the reader.]]** However, once correcting the RVs from stellar activity signal using a linear combination of the CCF asymmetry and width, the RV residuals obtained from a Normal or SN fitting are very similar. The precision on the asymmetric parameter derived from the SN is also 15% better than the one derived on the common Bisector Inverse Slope Span (BIS SPAN). Besides all these advantages, the errors on the RVs measured by the SN are 60% greater. **[[Referee: To me this statement alone tells me that the SN method is not very useful, and not one that I would use. Any method that produces larger errors *before* correcting for the activity is not a very useful method. Why should the reader continue beyond this point?]]** This is not a problem at signal-to-noise ratio levels > 150 because other noise sources dominate in this regime, however we have to be careful when deriving precise RVs for low SNR measurements. We also found that in the case of low SNR measurements, errors measured by bootstrap analysis can be 50% more precise than errors measured using the derivative

of the CCF. We strongly encourage the use of the SN distribution to derive the different moments of the CCF, because the derived moments probe better stellar activity signals than when using a Normal distribution. [[**Referee:** You simply cannot make this claim. What you have shown is that the SN method produces errors that are 60% larger than the Normal distribution and you hint that this is due to it being more sensitive to activity. You have not demonstrated this by looking at the "activity" signal that the SN isolates to see if it is actually more sensitive at finding rotation periods, activity cycles, etc. Where is the sensitivity in actually getting out useful information? Furthermore, after correcting the SN RVs for the activity, you arrive at the same answer as the Normal distribution method. I would conclude that the SN offers no real advantages over Normal so the community can continue to use the standard method. Grammatically this is an ambiguous statement. Are you looking for "better signals", i.e. ones that you like, or probing stellar signals in a better way? (This is what you meant, but not what was stated.)]]

1 Introduction

The radial velocity (RV) of a star is defined to be the velocity of the center of mass of the star along our line of sight. [[**Referee:** No, the radial velocity you measure of the star has many components that contribute to the RV: the space motion (present even if there is no companion), the motion about the center of mass (the so-called ?-velocity), oscillations, convective blue shift, etc. You are just assuming that you are measuring the reflex motion of a star-planet system. There are many physical processes that produce a Doppler shift besides exoplanets.]] This quantity can be derived precisely by measuring the Doppler shift of spectral lines produced in stellar atmospheres. [[**Referee:** Well, not if you have a poor measurement precision. You only get a "precise" measurement with simultaneous wavelength calibration. Again, the authors are not being very "precise" here.]] For spectrographs that are not stabilised in temperature and pressure, the iodine technique is used, where the light of the star passes through a iodine cell before getting into the spectrograph to imprint the absorption spectrum of iodine on top of the stellar spectrum (The Hamilton spectrograph (?) at Lick Observatory, HIRES (?) on the Keck 10-m telescope, the Tull spectrograph (?), the High Resolution Spectrograph HRS (?)).

In this case, if the spectrograph shifts due to changing atmospheric conditions, the iodine and stellar spectra are shifted in the same way. This leads to some complications when reducing the data because one has to decorrelate the iodine spectrum from the stellar spectrum. [[**Referee:** This is a poorly expressed thought that is technically wrong. The spectrograph does not shift due to changing atmospheric conditions (earth's atmosphere?). It changes because of mechanical shifts (vibrations, temperature changes in spectrograph housing, changes in instrumental profile) as well as movement of the image on the slit or fiber which yes, does affect stellar and iodine lines in the same way.]] [[**Referee:** No, you do not decorrelate (and I am not sure what the authors mean by this word) the iodine from the stellar spectrum. The authors show that they have no understanding how the method works. What is actually done is that a high-resolution iodine spectrum is combined with a spectrum of the star without iodine lines and a fit is made to the observed star+iodine spectrum.

It is not decorrelation, but rather χ^2 fitting.]

For spectrographs that are stabilised, the spectrum of a calibration lamp is recorded close to the stellar spectrum on the CCD, which prevents contamination of the stellar spectrum (CORALIE (?), The High Accuracy Radial Velocity Planet Searcher (HARPS) (?), HARPS- N (?), SOPHIE (?), CARMENES (?)). For those instruments, reducing the data is easier as the stellar spectrum is not contaminated with iodine absorption lines. **[Referee: This is not true! I have taken spectra with HARPS and there is cross-talk between spectra and Th-Ar particularly for strong emission lines. So there is some contamination. Furthermore, this "preventing of contamination" as the authors state has nothing to do with the spectrograph being stabilized, it has to do with the simultaneous wavelength calibration using two fibers. In fact one reference the authors give, CORALIE, is not a stabilized spectrograph!]** **[Referee: In paragraph 1 you give two lists of spectrographs. I presume the first one is a list spectrographs that use the iodine method, although that is not explicitly expressed. The authors talk about the iodine cell and then just give a parenthetical list of spectrographs. There is no explicit statement that these spectrographs, designed for other purposes, are all equipped with an iodine cell. The reader must assume this. Furthermore, the list is not all-inclusive. I can think of a dozen facilities that use iodine cells. So you should say, "For example" or..."to name a few" The second list it is for simultaneous Th-Ar calibration. Not all of these are "stabilized" like the authors claim. CORALIE, if I am not mistaken, was simultaneous Th-Ar without stabilization. That is why they built HARPS. Furthermore, all of these spectrographs were designed for precise stellar RVs, unlike the spectrographs listed for the iodine method.]**

For stabilized spectrographs, the RV is derived by first correlating the stellar spectrum with a synthetic (??) or an observed stellar template (?), which gives an average line profile, generally called Cross Correlation Function (CCF). Then a Normal distribution is fitted to this average profile to get the mean, namely the RV, and the full width at half maximum (FWHM) of the profile. The CCF technique allows for an averaging out of the RV information of thousands of lines and therefore reach a very high signal-to-noise ratio (SNR), which is essential for a good RV precision.

The convection in external layers of solar type stars is responsible for the granulation pattern than can be seen at high spatial resolution on the surface of the Sun. The differences in flux and velocity between upflows and downflows change the Normal profile of spectral lines that become asymmetric with a "C"-shaped profile (?). **[Referee: Regarding the C-shape of the bisector. The classic C-shape is only for late-type stars. In fact for hotter stars (F-type) it reverses. The strength of the asymmetry depends not just on the velocities, it is more complicated than that. It also depends on the flux ratio between hot cells and cool lanes, as well as the ratio of surface areas between the two.]** The strength of the asymmetry depends on the velocity of the convection, approximatively 300 m s^{-1} for the Sun, but also on the formation depth of spectral lines (?). Since the CCF is an average of all the spectral lines, some strongly asymmetric and some not, its asymmetry is rather small, which is why a Normal distribution is a proper model to fit the CCF. This small asymmetry modifies however slightly the estimated RVs of the star, reducing the accuracy of the measurement, but if this asymmetry does not vary with time, the precision is kept. **[Referee: But the activity causes the asymmetry to change due to rotation, spot evolution, etc. The authors are confusing the readers here. They talk about a changing activity signal, but now argue that the precision is**

only kept if the asymmetry is constant, which it is not. I think what they want to say is that you measure a precise stellar radial velocity, but the activity makes its own contribution which reduces the accuracy of the RV determination for the barycentric motion due to a planet.]]

Convection is not the only phenomenon responsible for asymmetries in the single spectral line and the CCF. Stellar activity is responsible for the appearance of dark spots and bright faculae on the stellar photosphere, which breaks the flux balance between the red-shifted and the blue-shifted halves of a rotating star and therefore induce an asymmetry of spectral lines and thus of the CCF. As the star rotates, spots and faculae move across the stellar disk, modifying the line asymmetry and thus producing an apparent Doppler shift (??????). Spots and faculae are also regions where the magnetic field is strong. Strong magnetic fields reduce stellar convection, which in turn modifies the asymmetry of spectral lines (??????).

Stellar activity induces RV variations by a modification of the spectral line asymmetry, while an orbiting companion only induces a pure Doppler shift on spectral lines without modifying their shape. Therefore, assuming that there is no instrumental systematics, stellar activity will induce a variation in line asymmetry or FWHM of the CCF. The line asymmetry is commonly retrieved by calculating the bisector of the CCF (?) and deriving the bisector curvature (?) or the bisector inverse slope span (BIS SPAN, ?).

? proposed different indicators, including a bi-Gaussian fitting of the CCF. Unfortunately, when analyzing slow rotators stars such as the Sun, due to the limited spectral resolution of the spectrographs and the limited precision in RV, it becomes difficult to measure the line asymmetry, resulting in complications for detecting very small-mass planets with the RV technique. [[Referee: It is the slope or span, not both. And you generally do not calculate the inverse, just the span. This may be correlated or inversely correlated with the RV.]] In the procedure described above, the measurement of the RV and the FWHM is done separately from the measurement of the line asymmetry. All those parameters are correlated when stellar activity is dominant, and performing a step-by-step approach makes it difficult to correctly derive the errors on the different parameters retrieved. [[Referee: I simply do not understand this statement, possibly because it does not follow a standard English construction. You have a CCF and you measure a width and bisector span. This is a measurement error determined by the S/N, the resolution, and the stability of the spectrograph. You can determine an error in a quantity (RV and BIS). You can also have an uncertainty in the contribution of activity to the RV signal, but that is not a formal error. True errors come from photon statistics, systematic errors, instrumental errors, etc. You can attempt to remove the signal due to activity and that has an associated error, but that error is due to your lack of knowledge of the surface structure causing this RV and the exact value of the contribution to the RV. This is different from a measurement error. Maybe that is what you mean here, but it is not clear from what is written.]] In addition, the Normal distribution cannot take into account the natural asymmetry of the CCF, leaving correlated noise in the residuals, which also complicates the determination of errors. We propose to overcome these problems by fitting a SN (SN) distribution to the CCF, which naturally includes a skewness parameter (?).

The paper is organized as follow. In Sec. 2 we introduce the SN distribution and motivate its use for fitting the CCF. In Sec. 3 we show that the SN distribution is a better representation of

observed CCF and we study how the SN parameters relate to the RV, FWHM and BIS SPAN of the CCF. In Sec. 4 we present a simple model to correct for stellar activity. In Sec. 6, we compare on real observations the sensitivity of the SN parameters to stellar activity with respect to other existing indicators. In Sec. 7 we derive error bars for the different CCF parameters, and finally we discuss our results and conclude in Sec. 8 and Sec. 9.

2 The Skew Normal distribution

Jessi: This section is too complicated. I suggest we define the skew normal in terms of the parameters that are used later, then relegate additional details, including the optimization, to an appendix section.]

Referee: 17) I cannot find where R, R+, x and w were defined. It is a common practice when you introduce variables in a paper, these should all be defined. 18) Paragraph 4: "By using the so-called direct parameterization (DP) defined in Eq. 2 problems with parameters estimation arise in the evaluation of the likelihood function in a neighborhood of $\alpha = 0$, where the presence of a stationary point complicates the achievement of the maximum likelihood or least squares estimates." A poorly constructed sentence that simply confuses me! The authors are trying to say too many things in one sentence which loses the reader. 19) To be honest, I could not follow all the mathematics. I really would have to read the referenced material, but I had little time. Maybe the details can be put in an appendix?]

The Skew Normal (SN) distribution, introduced by ? **Jessi:** Wikipedia says this was introduced in O'Hagan and Leonard (1976), Bayes estimation subject to uncertainty about parameter constraints. However, it seems this may just be a 2-parameter class. I don't know that we need to get into the history anyway, we can just discuss the SN and reference the Azzalini paper.] , is a class of probability distributions which includes the Normal distribution as a special case. The SN distribution has, in addition to a position and a scale parameter, a third parameter which describes the asymmetry of the distribution.

Formally the density function of the SN distribution for a random variable $Z \in \mathbb{R}$ is defined as:

$$SN(z; \alpha) = 2\phi(z)\Phi(\alpha z), \quad (1)$$

where ϕ and Φ are respectively the density function and the distribution function of a standard Normal distribution and $\alpha \in \mathbb{R}$ is the skewness parameter which quantifies the asymmetry of the SN. It is possible to provide a general expression where both the location and the scale parameters are included. If $Z \sim SN(\alpha)$, $\xi \in \mathbb{R}$ and $\omega \in \mathbb{R}^+$, the density of $Y = \xi + \omega Z$ is defined as:

$$SN(z; \xi, \omega, \alpha) = \frac{2}{\omega} \phi\left(\frac{y - \xi}{\omega}\right) \Phi\left(\alpha \frac{y - \xi}{\omega}\right) \quad (2)$$

and $Y \sim SN(\xi, \omega^2, \alpha)$.

From both a mathematical and an inferential viewpoint, the SN distribution has useful properties; the details can be found in ?. As previously mentioned, the Normal distribution is included

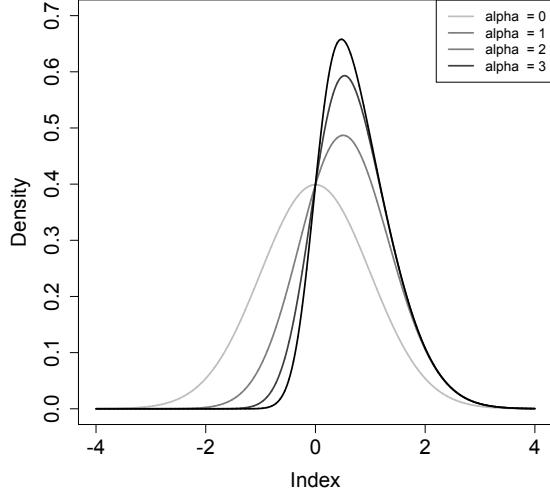


Figure 1: Density function of a $SN(\alpha)$ for different choices of α . For negative values of α the asymmetry will be on the left tail of the distribution.

in the SN family when $\alpha = 0$. The parameter α provides information about the asymmetry of the curve. In Figure 1, we plot the density function of the SN distribution for different values of α .

By using the so-called direct parameterization (DP) defined in Eq. 2 problems with parameters estimation arise in the evaluation of the likelihood function in a neighborhood of $\alpha = 0$, where the presence of a stationary point complicates the achievement of the maximum likelihood or least squares estimates. In order to solve this problem, ? proposed to pass from the direct parameterization having parameters ξ, ω and α to the so-called centred parameterization (CP), which is introduced below.

According to Eq. 2, $Y = \xi + \omega Z = \mu + \sigma Z_0$, where $Z \sim SN(\alpha)$ and Z_0 is defined below. Hence the first two moments of Z are defined simply with:

$$\begin{aligned}\mu_z &= E(Z) = b\delta, \\ \sigma_z &= var(Z) = 1 - (b\delta)^2,\end{aligned}$$

where $b = \sqrt{2/\pi}$, $\delta = \alpha/\sqrt{1+\alpha^2}$ and Z_0 , which is the standardization of Z , is defined as:

$$Z_0 = \frac{1}{\sigma_z} (Z - \mu_z). \quad (3)$$

By using Eq. 3, the new set of parameters (μ, σ^2, γ) provides a more clear interpretation of the behavior of the SN distribution. In fact, as it happens in the Normal case, μ and σ^2 indicate

exactly the mean and the variance of the distribution, while γ naturally represents an index for evaluating the asymmetry of the SN, making its interpretation simpler than the one available when α is considered. In order to completely define the CP, we need to express the CP parameters (μ, σ^2, γ) as a function of the one used in the DP (ξ, ω^2, α) :

$$\begin{aligned}\mu &= E(Y) = \xi + \omega\mu_z, \\ \sigma^2 &= var(Y) = \omega^2(1 - \mu_z)^2, \\ \gamma &= \frac{4 - \pi}{2} \frac{\mu_z^3}{(1 - \mu_z^2)^{3/2}}.\end{aligned}\tag{4}$$

By simple algebra, we can invert those equations to get the DP parameters as a function of the CP ones.

$$\begin{aligned}\xi &= \mu - b\omega\delta, \\ \omega &= \frac{\sigma}{\sqrt{1 - b^2\delta^2}}, \\ \alpha &= \frac{R}{\sqrt{b^2 - (1 - b^2)R^2}}, R = sign(\gamma) \left[\frac{2|\gamma|}{4 - \pi} \right]^{\frac{1}{3}}.\end{aligned}\tag{5}$$

The CP allows us to use the usual methods offered by the inferential theory and we will use this parameterization for the rest of the paper. Further details about DP, CP and statistical properties of the SN are treated in rigorous mathematical and statistical viewpoints in the book by ?.

3 Fitting the Skew Normal distribution to the CCF

We fit the following function to the CCF:

$$C - A \times SN(y; \mu, \sigma^2, \gamma),\tag{6}$$

where C is a constant fitting the continuum of the CCF, A is an amplitude parameter and y is the RV of each point of the CCF. Note that the CCF is expressed in flux as a function of the lag of the cross-correlation template, expressed in RV.

We use least squares to fit the SN distribution to the CCF. Since each point of the CCF has a measurement error, we include this information in the specification of the SN distribution by introducing the following heteroskedastic function for the variance:

$$var(Y_i) = \sigma^2 + \sigma_i^2, i = 1, \dots, n,\tag{7}$$

where n corresponds to the number of points of the CCF and σ_i represents the measurement error for each available point. By using Eq. 7 we can interpret the parameter σ^2 as a measure of the pure variability present in the evaluation of the CCF. However, although useful for estimating correct error bars in the case where the different σ_i do not include all systematic, this heteroskedastic

function only consider Gaussian distributed residuals. As the CCF is the result of a cross correlation, we expect the residuals of a Normal or a SN fit to show some red noise. For this reason, we first calculated the best fit for each CCF using a Normal or a SN distribution assuming Gaussian distributed noise. We then subtracted the median of the residuals obtained for all CCFs before performing a second fitting round. The result of this procedure is illustrated in Fig. 2. We see in the lower left panel that the residuals of a simple fit of a Normal or a SN distribution are strongly correlated. After removing the median of the residuals, in the lower right panel, the final residuals of the SN have nearly no correlation, while the final residuals of the Normal still present a small systematic component. This comes from the fact that because of the asymmetric nature of the CCF (see Sec. 1), the SN distribution is a more flexible model than the Normal one.

When using the least squares estimate for fitting the CCF with the SN distribution, we have to take into account the stationary problem in $\gamma = 0$ due to the presence of the function $\text{sign}(\gamma)$ in Equation (5). Depending on the initial values of the numerical minimization for the least squares estimate of γ , the presence of this stationary point when evaluating γ can lead to an incorrect estimation of the asymmetry parameter of the SN. In order to resolve this problem, we implemented two functions for retrieving the best least squares solution. In the first function we allowed for only positive solutions of γ_1 (with an upper limit of 0.995), while in the second function only negative solutions are allowed (with a lower limit of $\gamma = -0.995$). The chosen solution is the estimate which has smallest residuals.

In the following of the paper, we define the RV as the mean of the Normal or the mean of the SN (defined as SN RV). For the width of the CCF, we use the FWHM of the Normal, defined as $2\sqrt{2 \ln 2}\sigma$. The width of the SN, SN FWHM, is defined in the same way¹. In the Normal case, it is not possible to retrieve an asymmetric parameter and therefore we use BIS SPAN. This will be compared to the asymmetric parameter γ_1 of the SN.

4 Effect of stellar activity on radial velocities and correction

Exoplanets will only produce a RV variation induced by a pure Doppler shift of stellar spectra. Stellar activity, on the contrary, does not produce a blueshift or redshift of the spectra, but creates a spurious RV signal by modifying the shape of spectral lines. To track these changes in line shape, we generally use the FWHM, the BIS SPAN or the indicators introduced by ? that give an information on the average width and asymmetry of the CCF. A strong correlation between the RVs and one or more of these parameters is a sign that the RVs are significantly affected by stellar activity signals.

To test the strength of the correlation between RVs and the different activity indicators, we calculated the Pearson correlation coefficient defined as:

$$R(x, y) = \frac{\text{cov}(x, y)}{\sigma(x), \sigma(y)}, \quad (8)$$

¹Note that SN FWHM does not correspond to the width of the SN distribution at half maximum like in the Normal case.

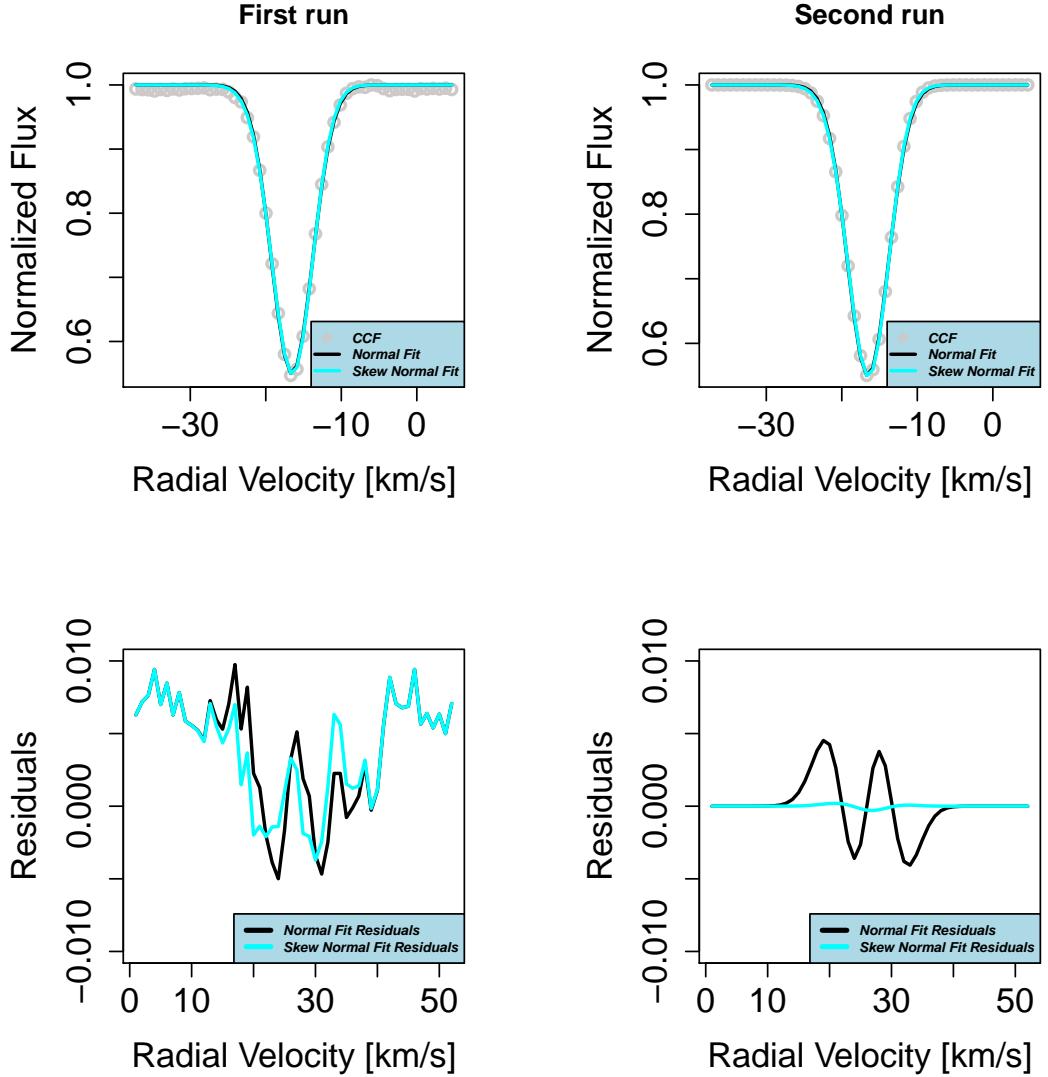


Figure 2: Comparison between the Normal and the Skew Normal fit for a particular CCF for the star Tau Ceti. The first run corresponds to fitting a Normal or a SN to the raw CCF. The second run corresponds to fitting the same function to the CCF after removing the median of all the residuals obtained in the first run. The SN residuals of the second run present a tiny correlation, whereas the Normal residuals present still a clear systematic component.

where x and y are two quantitative variables, $\text{cov}(x, y)$ indicates the covariance between x and y , and $\sigma(x)$ and $\sigma(y)$ represents their standard deviation. A p -value for the statistical test having null hypothesis $H_0 : R = 0$ is provided, along with a 95 % confidence interval for R .

To correct for stellar activity signals, it is common to consider a linear combination of the RVs with the BIS SPAN and the FWHM (or γ and SN FWHM in the SN case): [[Referee: Is this really common? I am not aware of it having wide use in the community. If so, references would support this. The authors have to distinguish what their research group and collaborators do from what the rest of the community does.]]

$$RV = \beta_0 + \beta_1 \gamma + \beta_2 FWHM + \epsilon, \quad (9)$$

where β_0 is the intercept and ϵ is the vector of the errors with mean equal to 0 and covariance matrix equal to $\sigma^2 I$ (I is the identity matrix). [[Referee: Is this expression just for the RV from activity alone, or the measured RV from the star (contribution of all phenomena)? If it is just activity, you should put a subscript otherwise the readers will be confused. In the text you refer to the variable ?1, but Eq. 9 has no subscript in ?. You give one equation for Normal and SN, but in the text you state these variables have to be replaced by such and such when using SN, etc. It will be less confusing if you simply have two equations, one for Normal and one for SN and with their respective variables.]] When the Normal fit is used, the parameter γ_1 is replaced by the BIS SPAN and SN FWHM by FWHM. In order to show the goodness of this correction, a statistical test on the parameters β_0 , β_1 and β_2 is presented, where the null hypothesis is $H_0 : \beta_i = 0$, for $i = 0, 1, 2$. The level for not rejecting the null hypothesis is fixed equal to 0.05. Pearson correlation coefficient R^2 is considered for explaining how well this linear combination explains the variability of the RVs of the star due to stellar activity.

5 Simulated data examples

5.1 Doppler shift added to Alpha Centauri B

6 Real data applications

[[Jessi: I suggest we include one real-date application (whichever is used in §5.1 (Alpha Centauri B?) and then add the other examples to an appendix]]

In this Section we present the analyses of five stars, showing the advantages of fitting the CCF using the SN distribution defined in Section 3.

A comparison with the results obtained by the classic approach is done, where the RVs of the star are estimated by the mean of the Normal distribution used to fit the CCF along with the BIS SPAN or the other asymmetric parameters defined in ?.

6.1 Alpha Centauri B

We first analyze Alpha Centauri B, where 1816 CCFs in 2010 are considered. Several measurement in 2010 are contaminated by light from Alpha Centauri A. To remove contaminated spectra and thus CCFs, we performed the same selection as presented in ?.

We begin by evaluating the correlation between γ and the classic BIS SPAN. In Fig. 3, we see that the relationship between γ and the BIS SPAN is linear, with a slope equal to 0.933 and a strong Pearson correlation coefficient of 0.958.

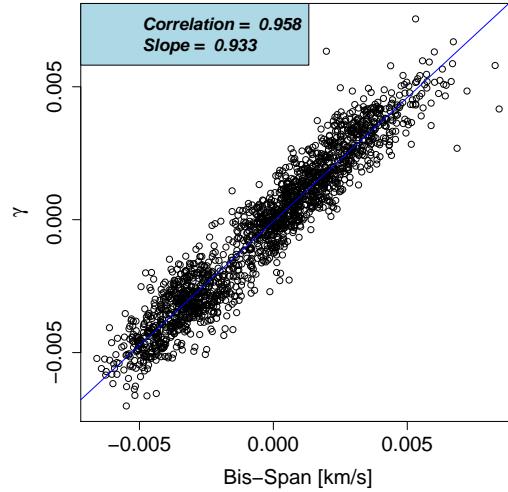


Figure 3: Correlation between γ and the BIS SPAN for Alpha Centauri B.

Figure 4 shows the comparison between the RVs retrieved using the SN distribution and the ones obtained with the Normal distribution. This data set for Alpha Centauri B, used in ? and ?, present a strong stellar activity signal. We see that the RVs measured by the SN fitting show more variations than the RVs measured by the Normal fitting. [[Referee: This is not at all clear from the figure. Is there a way to quantify this?]] The SN fit is therefore significantly more sensitive to stellar activity. [[Referee: How can you be sure that the higher scatter in your measurement error is not simply due to the fact that you have an inferior measurement technique? This can be answered by isolating the RV contribution from activity (from the asymmetry). Is this random (noise) or is there structure (activity).]] [[Referee: If you are interested in using the RV signal due to activity to measure the rotation period, effects of activity, etc., why not just show the RVs due to the activity signal alone.?]] This can be explained by the fact that because the SN includes an asymmetry parameter, the RV defined as the mean of the SN distribution gets more shifted in the direction of the asymmetry induced by stellar activity.

We show the correction of stellar activity using Eq. 9 in Fig. 5. We see that after correcting for

| Parameter | Normal Fitting | Skew Normal Fitting |
|-----------|----------------|---------------------|
| β_0 | 0.0066 | $2.29e - 11$ |
| β_1 | 0.022 | $2.22e - 16$ |
| β_2 | $2.22e - 16$ | $2.22e - 16$ |
| R^2 | 0.49 | 0.77 |

Table 1: **Alpha Centauri B:** Evaluation of the linear combination used for correcting the RVs from stellar activity, according to Eq. 9. The p-values for the parameters β_0 , β_1 and β_2 for both the methodologies are summarized, as well as the R^2 . Although more significant for the SN, all parameters are useful in explaining the variability of the RVs of the star. The evaluation of the R^2 shows how the linear combination better explains the variability in RVs due to stellar activity coming from the SN analysis.

stellar activity, the rms of the RV residuals in the Normal and SN analysis are extremely similar. However, we note that when using the SN analysis, the correction is more important. This is confirmed by the statistical tests on the significance of the parameters β_0 , β_1 and β_2 whose results are summarized in Table 1. The intercept and both the variables γ and SN FWHM are necessary for correcting the RVs retrieved using a SN and Normal fitting. The comparison of R^2 shows as well that the SN case accounts for a higher percentage of variability in RVs.

We compare the correlation between the different activity indicators and the RVs of the star in Fig. 6. The correlation between γ_1 and RVs is much stronger, almost twice, than the correlation calculated between the other asymmetry parameters and their corresponding RVs. The correlation between FWHM and the RVs is as well stronger when fitting a SN distribution. All the correlations are statistically different from 0.

By looking closer to the data in Fig. 7, we saw that there are three distinct temporal clusters in the Alpha Centauri B measurements and those clusters have a different linear relationship between γ_1 and RVs (and also SN FWHM and RVs, though not displayed here). When considering Eq. 9 and the subsequent inferences, this clustering is not accounted for in the model. A slightly more general linear model that allows for different intercepts and different slopes for the three clusters for γ_1 and SN FWHM can be considered. Adjusting the RVs for stellar activity using this expanded model produces the corrected radial velocities displayed in the left plot of Fig. 8. Those corrected RVs are different from those displayed in the lower left plot of Fig. 5 (which apply the correction derived from Eq. 9); the right plot of Fig. 8 displays the difference between the two sets of corrected RVs. The long-term trend can be explained by the fact that the RV drift induced by the companion Alpha Centauri A is not well corrected, however the shorter-term variations show that stellar activity varies as a function of time, with spots and faculae evolving, and therefore modifies the dependence of the RVs with respect to γ_1 and SN FWHM. These temporal variations is something that we want to explore more, however this work presents a significant effort that we want to address in a future paper.

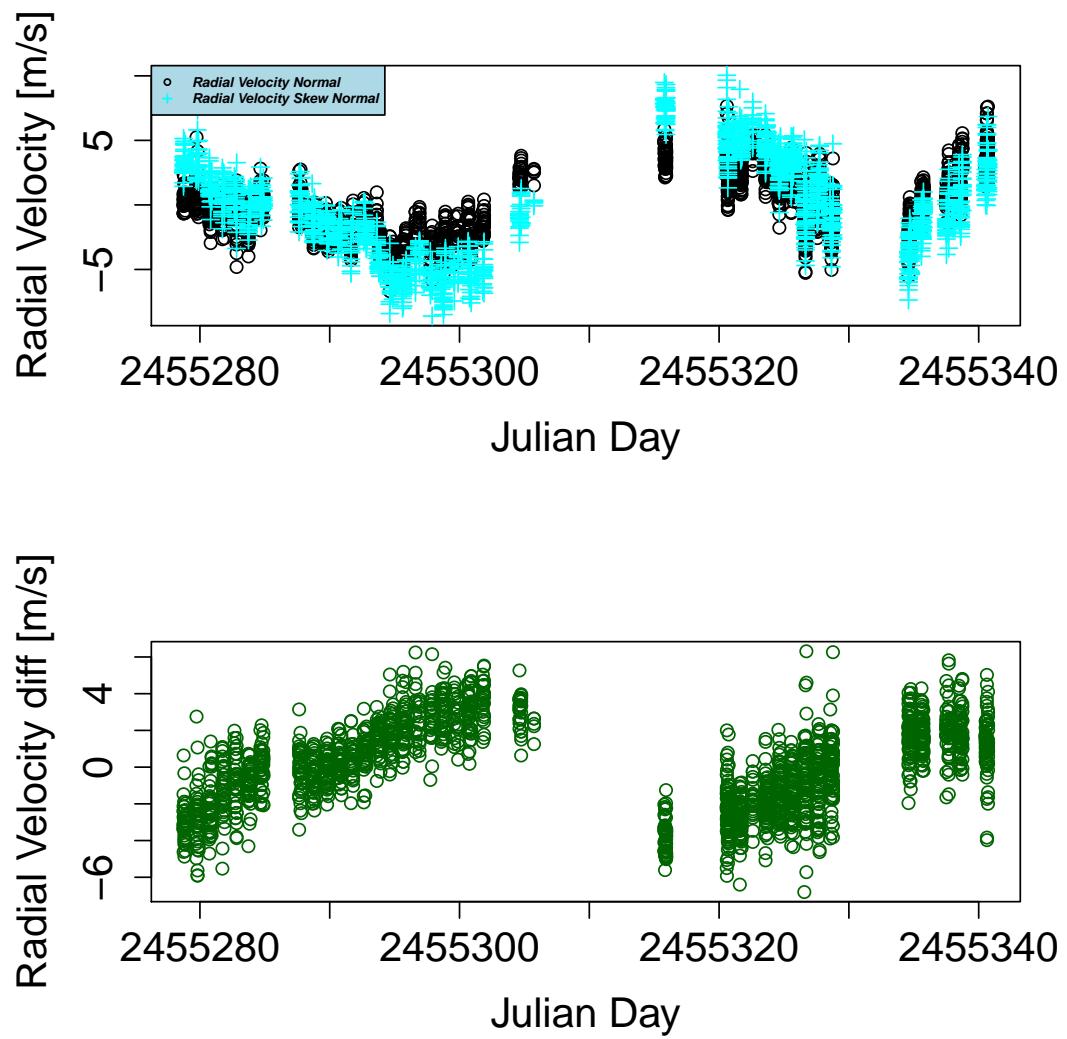


Figure 4: RVs and RV differences for Alpha Centauri B considering a Normal and a SN fitting.

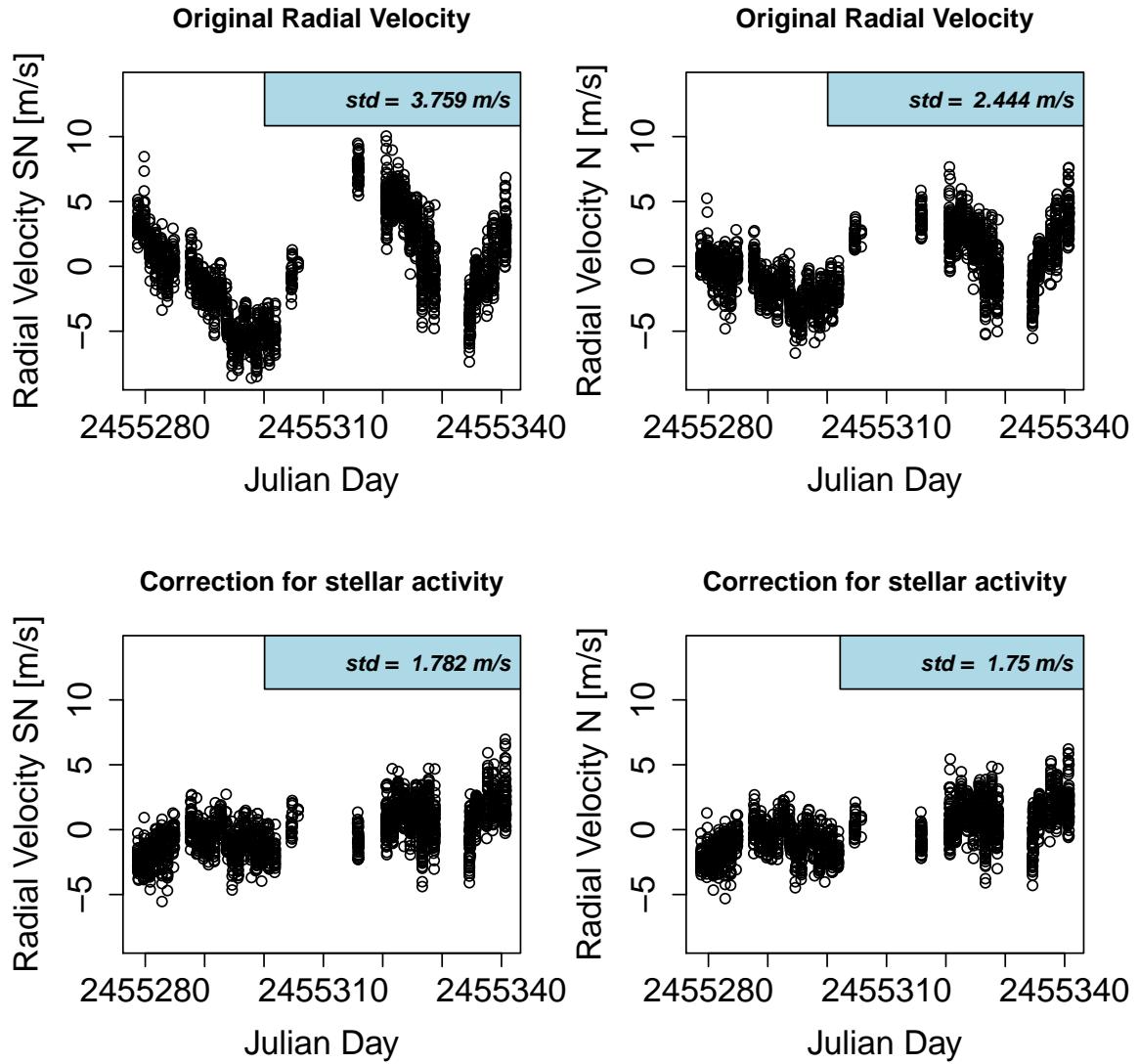


Figure 5: RVs of Alpha Centauri B using a Normal and a SN fitting before and after correcting for stellar activity using Eq. 9. [[Referee: Fig 5. The differences in the results of the two methods appear negligible to me]]

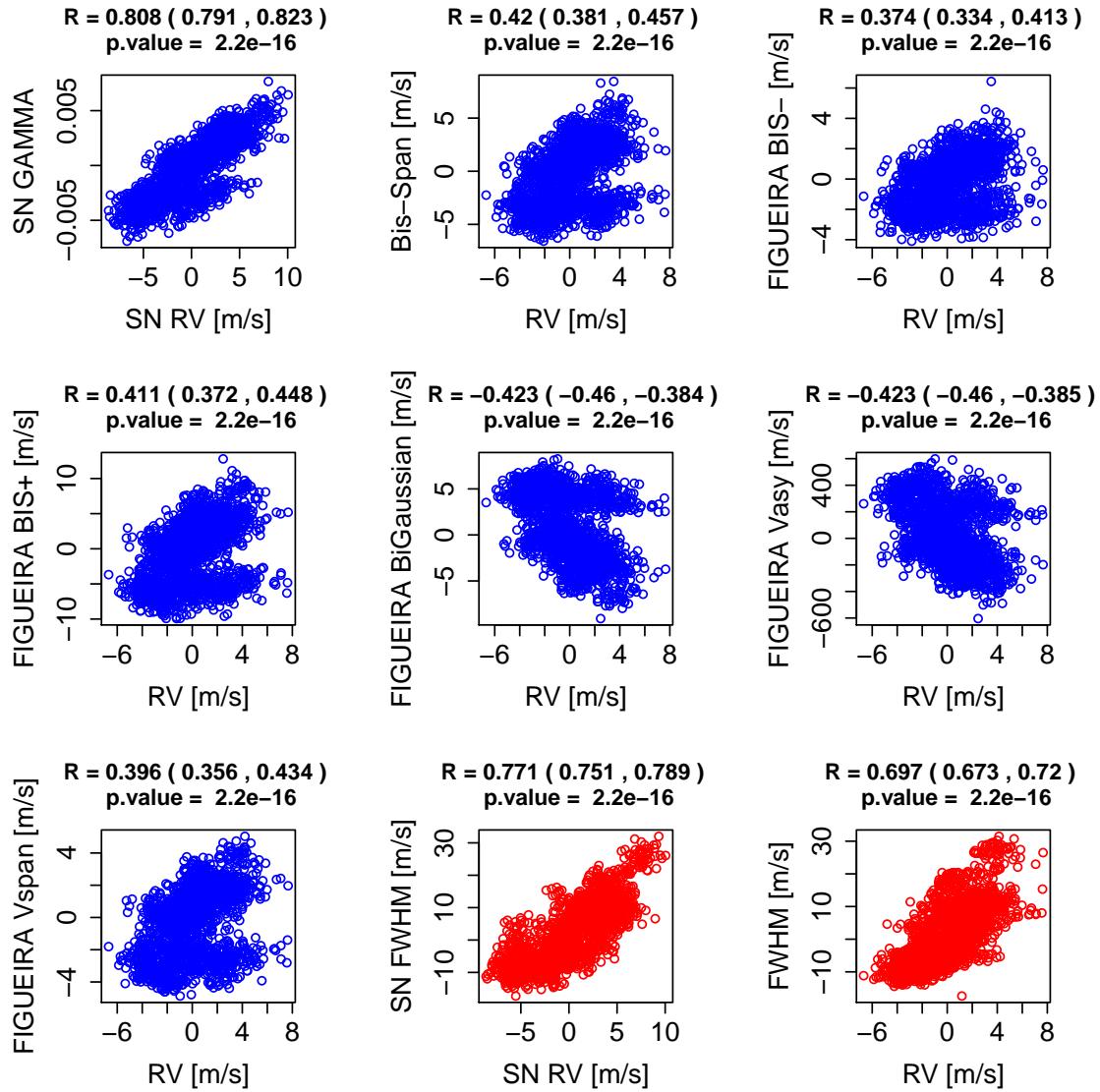


Figure 6: Correlation between the asymmetry parameters and the RVs for Alpha Centauri B. The last two plots show the correlation between the FWHM and the RVs for Alpha Centauri B using respectively the SN and the Normal fits.

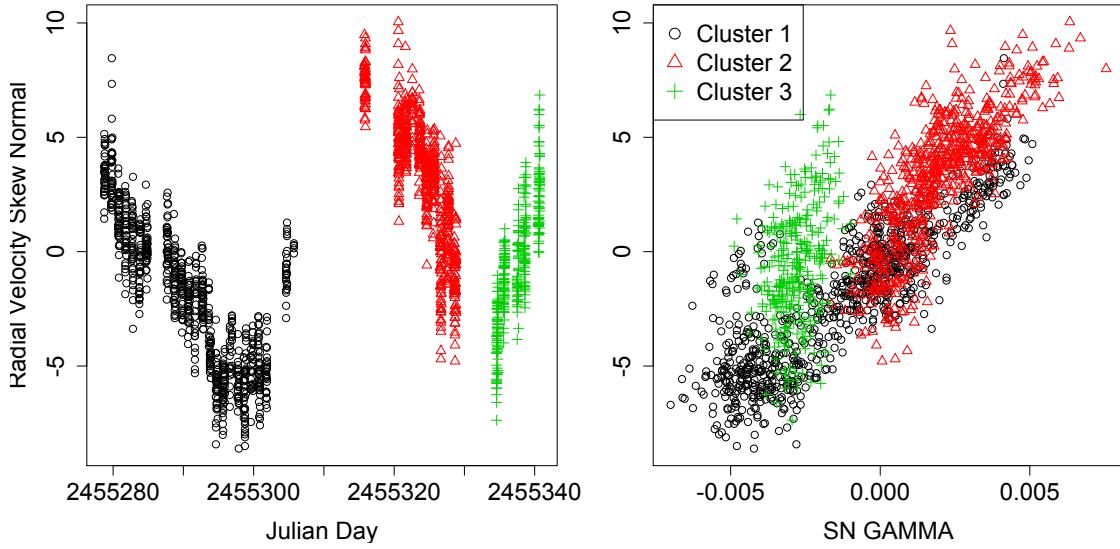


Figure 7: The RVs as a function of time (left) and the RVs plotted against γ_1 (right) with colors and plot symbol according to its temporal cluster assignment for Alpha Centauri B. The RVs are expressed in m s^{-1} .

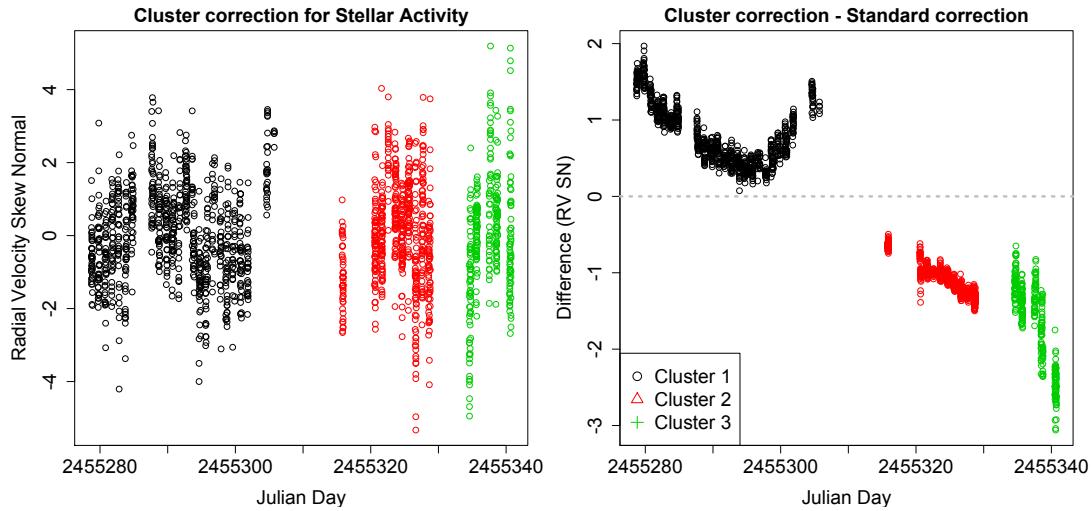


Figure 8: The RV for Alpha Centauri B corrected for stellar activity using the SN fit and accounting for the temporal clusters (left), and the difference between those values and the analogous values without accounting for the temporal clusters (right) which are displayed in the lower left plot of Figure 5. The RVs are expressed in m s^{-1} .

6.2 HD192310

We present now the results of the analysis for the star HD192310 (i.e. Gliese 785). The dataset consists in 1588 CCFs. Figure 9 shows the correlation between γ and BIS SPAN to understand the link between those two parameters. It is interesting to see that, in this case, the slope of 0.645 is significantly different from the one found for Alpha Centauri B in Fig. 3. [[Referee: Why is this interesting? You would expect it to be given that this is a different star! If something is interesting you have to point out to the reader why.]]

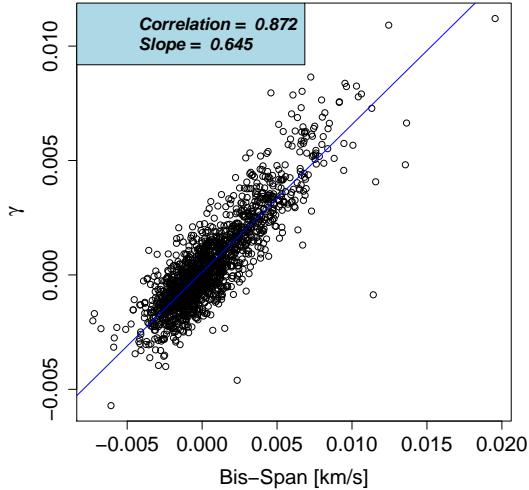


Figure 9: Correlation between γ and the BIS SPAN for HD192310.

In Fig. 10, we see like for Alpha Centauri B that the RVs obtained with the SN analysis present a larger rms than the RVs obtained with the Normal fitting. However, once correcting for stellar activity using a linear correlation with γ and SN FWHM (or BIS SPAN and FWHM, see Eq. 9), the rms of the residuals are comparable. We note that when correcting for stellar activity, the BIS SPAN is not statistically significant (see Table 2). Like for Alpha Centauri B, the Pearson correlation coefficient R^2 shows that the model we used to correct for stellar activity is a better one in the SN case than in the Normal case. [[Referee: So if I understand this correctly, SN gives higher scatter in the RV, but you make a bigger correction to the activity. In the end you arrive at the same answer as the Normal distribution and using Eq. 9 . So why use SN?]]

The comparison between the asymmetry parameters and the RVs are presented in Figure 11. The γ parameter shows a stronger correlation with the RVs (0.587) than the other asymmetric parameters. The correlation between the CCF width and the RVs is also stronger when fitting a SN rather than a Normal.

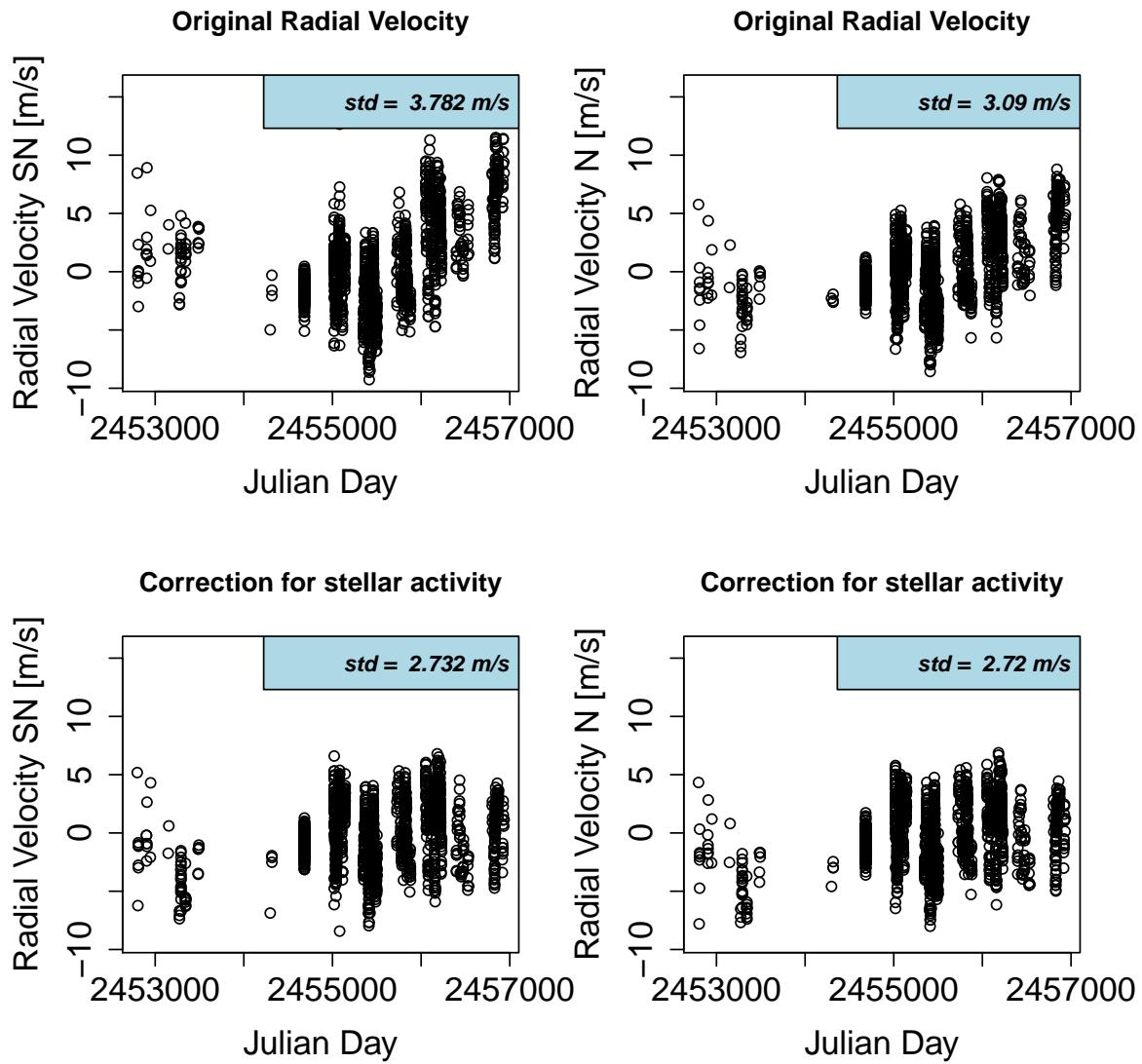


Figure 10: Radial Velocities for HD192310 using the Normal and the Skew Normal fitting before and after correcting the analyses from the stellar activity.

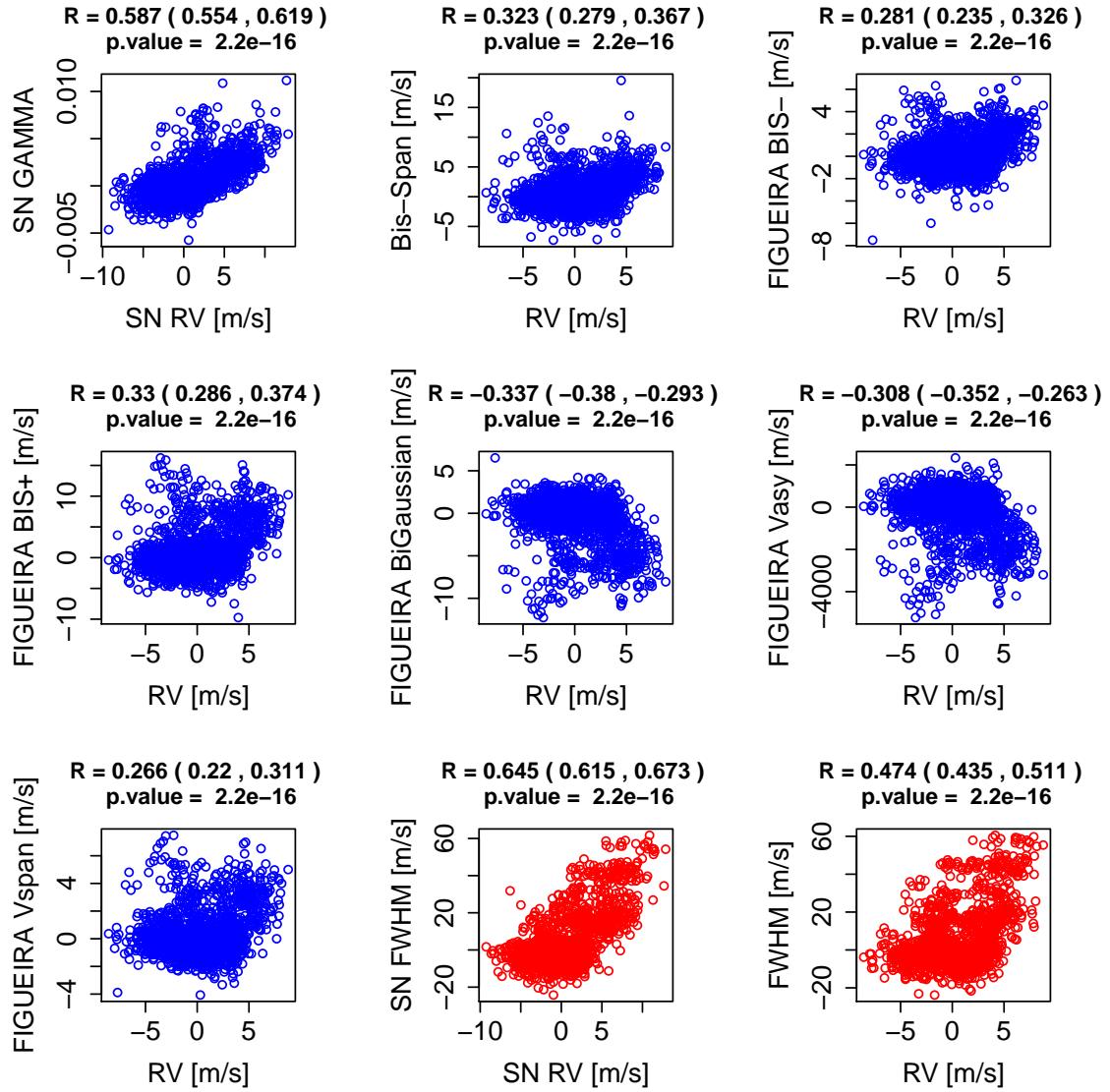


Figure 11: Correlation between the asymmetry parameters and the RVs for HD 192310. The last two plots show the correlation between the FWHM and the RVs for HD 192310 using respectively the SN and the Normal fits.

| Parameter | Normal Fitting | Skew Normal Fitting |
|-----------|----------------|---------------------|
| β_0 | $2e - 16$ | $2.22e - 16$ |
| β_1 | 0.18 | $2.22e - 16$ |
| β_2 | $2e - 10$ | $2.22e - 16$ |
| R^2 | 0.23 | 0.48 |

Table 2: **HD192310:** Evaluation of the linear combination used for correcting the RVs from stellar activity, according to Eq. 9. The p-values for the parameters β_0 , β_1 and β_2 for both the methodologies are summarized, as well as the R^2 . Concerning the Normal fitting, the BIS SPAN is not statistically useful in explaining the variability of the RVs of the star, whereas for the SN analyses all the variables are statistically different from 0. The evaluation of the R^2 shows how the linear combination better explains the variability in RVs due to stellar activity coming from the SN analysis.

6.3 Tau Ceti

[[Referee: If you are going to refer to the star as HD 10700 you should use the same name in the section heading.]] The analysis of the star HD10700 (i.e. Tau Ceti) consists of 7963 CCFs. Figure 12 shows the relationship between γ and the BIS SPAN, with a correlation of 0.605 and a slope of the fitted line of 0.566. Although still significant, we see that this correlation is weaker than in the case of Alpha Centauri B or HD192310. This is probably because HD10700 is at a very low activity level, similar to the Sun at its minimum activity phase. [[Referee: And how exactly do you arrive at this conclusion? It is ok to hypothesize that the star has a low activity level. But to claim it has the same activity level as the sun at solar minimum is completely unqualified and should be supported by evidence. This is especially true since HD 10700 is a late F-type star and thus may have a different activity behavior from a G2 star.]]

The RVs derived with the SN present a slightly more important rms as we can see in Fig. 13, and after correcting for stellar activity, the rms of the RV residuals are extremely similar. [[Referee: And just what is a "more important" rms compared to a less important one? The rms is a measurement of the properties of the data often with respect to a fit.]] In Table 3, we see that in the Normal and SN analysis, the intercept, the width and the asymmetry of the CCF (measured with FWHM or SN FWHM and γ or BIS SPAN) can explain part of the RV variability measured. Looking at R^2 , the RVs derived with the SN present a stronger relationship with activity parameters than the RVs derived in the Normal.

The evaluation of the correlation between the asymmetry parameters and the SN RVs shows how using the SN procedure leads to much stronger correlations. The correlation between γ and the RVs is 0.566, whereas the correlation between the common asymmetry statistics with their corresponding RVs result in values close to 0. The correlation between the FWHM and the RVs of HD10700 for the SN and the Normal are similar.

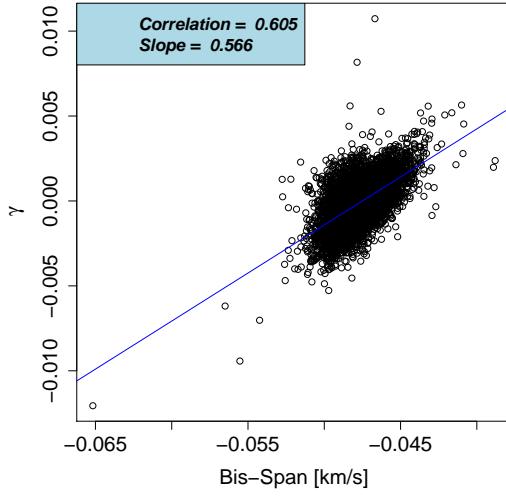


Figure 12: Correlation between γ and the BIS SPAN for Tau Ceti.

| Parameter | Normal Fitting | Skew Normal Fitting |
|-----------|----------------|---------------------|
| β_0 | 0.00013 | $2.22e - 16$ |
| β_1 | $4.83e - 6$ | $2.22e - 16$ |
| β_2 | $2.22e - 16$ | $2.22e - 16$ |
| R^2 | 0.27 | 0.43 |

Table 3: **HD10700**: Evaluation of the linear combination used for correcting the RVs from stellar activity, according to Eq. 9. The p-values for the parameters β_0 , β_1 and β_2 for both the methodologies are summarized, as well as the R^2 . For both the analyses, all the variables are helpful, although the linear combination is statistically more significant when working with the SN parameters.

6.4 HD215152

The analysis of the star HD215152 consists in 284 CCFs and Figure 15 shows that the slope of the linear regression between γ and the BIS SPAN is 0.591 with a Pearson correlation coefficient of 0.767.

In the case of HD215152, we see in Fig. 16 that the RVs measured with the SN or the Normal, and the corresponding RV residuals after activity correction all present very similar RV rms. Thus activity correction does not seem to be very efficient at understand the RV variation. [[Referee: Poor English. How can a correction be "efficient at understand [sic] a variation?" Revised construction: "Thus the activity correction does not seem to be useful in helping us understand the RV variations."]] This is probably because of the presence of planetary signals in the RV data (?).

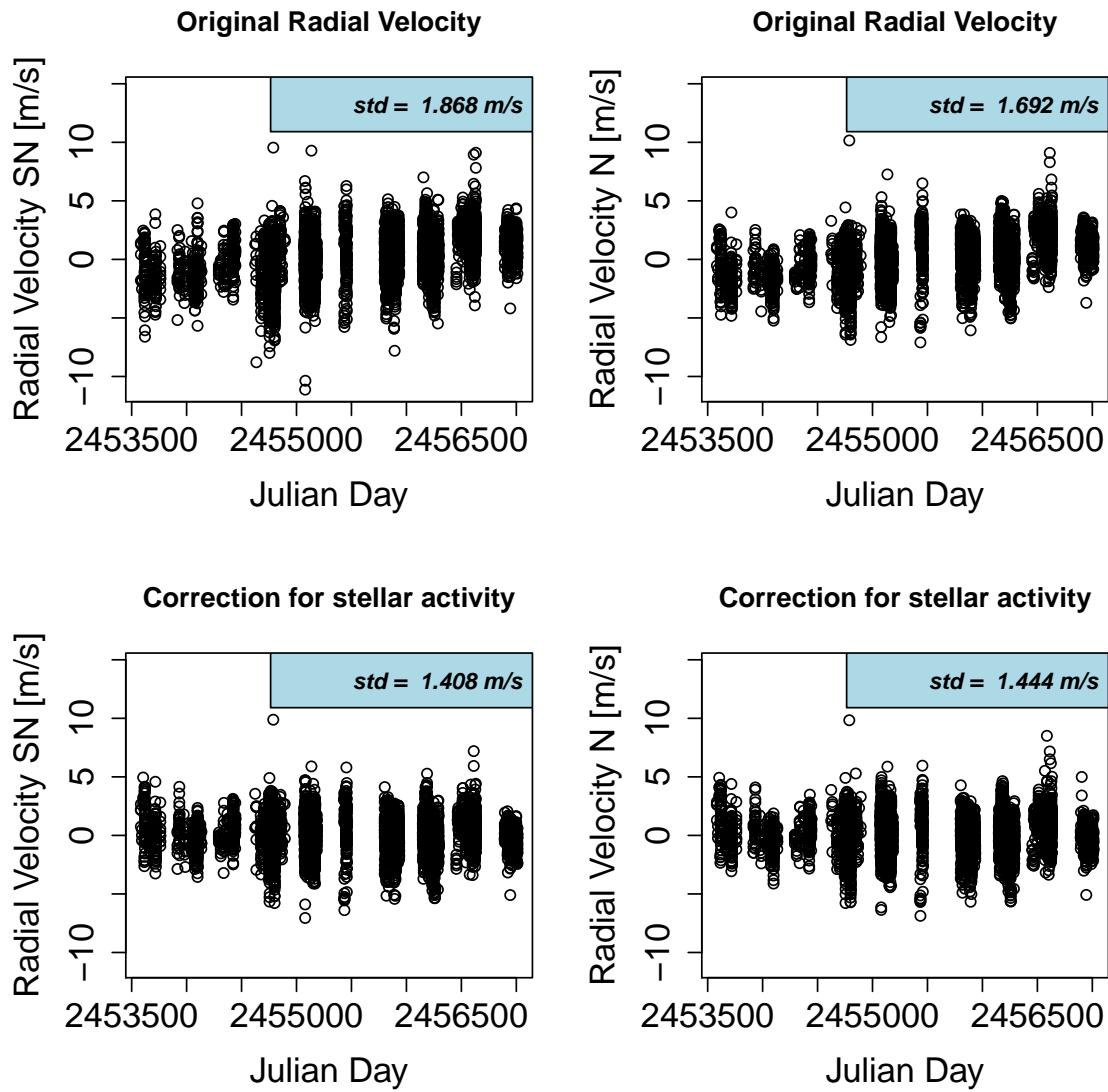


Figure 13: RVs for HD10700 using the Normal and the Skew Normal fitting before and after correcting for stellar activity.

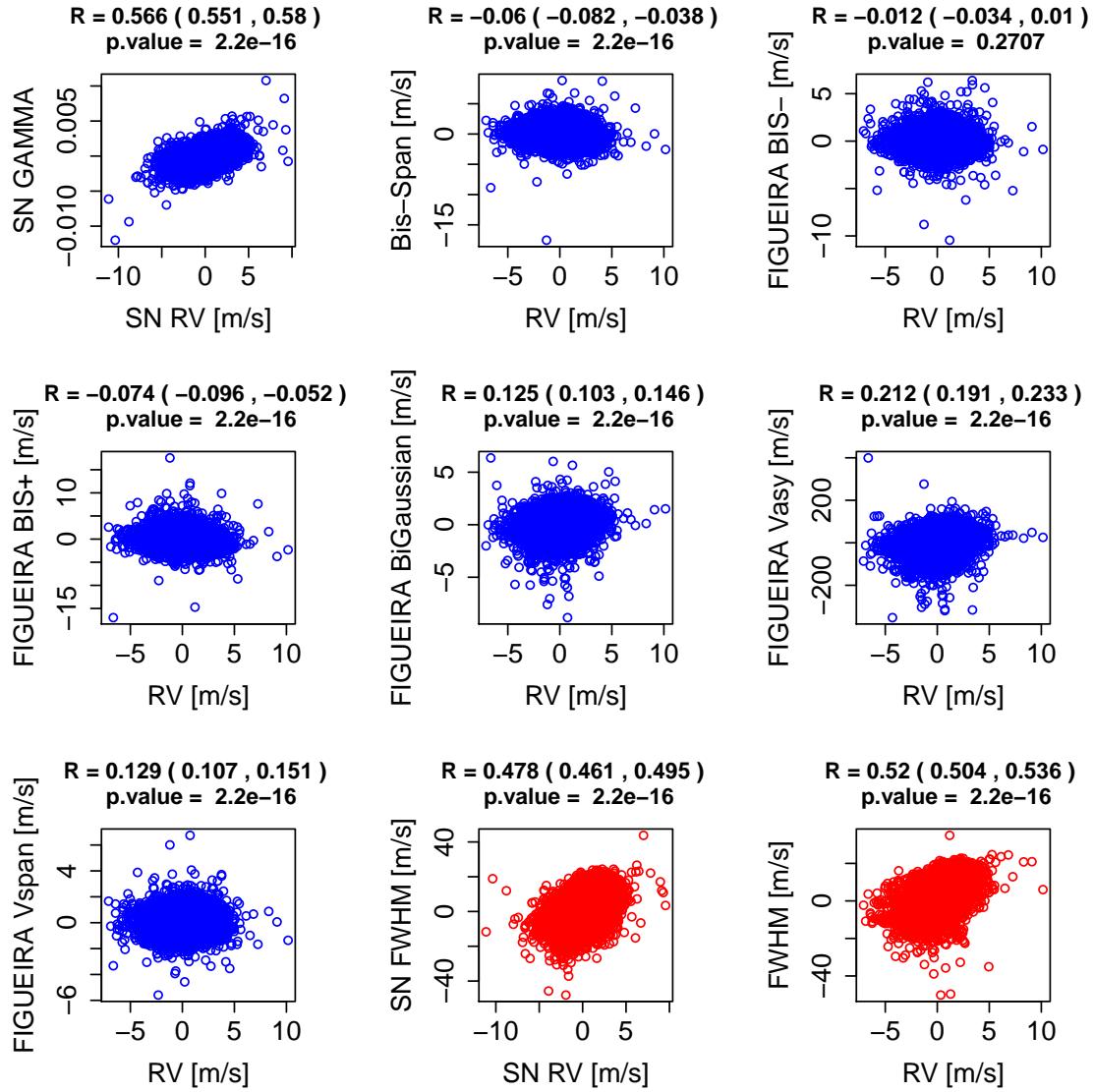


Figure 14: Correlation between the asymmetry parameters and the RVs for Tau Ceti. The last two plots show the correlation between the FWHM and the RVs for Tau Ceti using respectively the SN and the Normal fits.

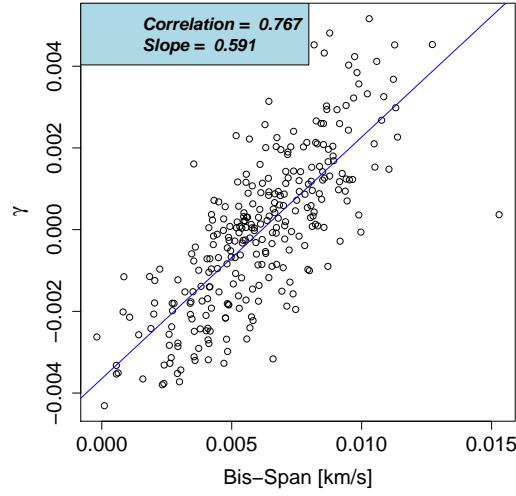


Figure 15: Correlation between γ and the BIS SPAN for HD 215152.

[[Referee: So there is a planetary signal in the data but you do not have an orbit that gives you a phase? Then this is not a confirmed planetary signal and thus cannot support your claim]] Note however that the information on the orbital phase of the planets is not present in ? and we cannot therefore remove those signals. The significance of β_0 , β_1 and β_2 in Table 4 shows that in the Normal case BIS SPAN is not significantly correlated with the RVs, while for the intercept and the FWHM the test with level 0.05 is barely significant. In the SN case it is the intercept and FWHM that are not significantly correlated. The Pearson correlation coefficient R^2 shows that the RV derived from the SN fit are more affected by stellar activity than the RVs measured with the Normal, although the correlation is low at a level of 0.15.

The correlations among the asymmetry parameters and the RVs shows also in this case that using the SN fitting provides a stronger result. The correlation between γ and the RVs is 0.38, whereas the correlation between the common asymmetry statistics and their corresponding radial velocities is smaller than 0.30. The SN FWHM or the FWHM are not significantly correlated to the SN RVs or the RVs.

6.5 Corot-7

The final star considered is Corot-7 which has low signal to noise. [[Referee: A star cannot have a "low signal to noise" only data has that. And what data are you referring to photometry (from CoRoT) RVs, from HARPS? You need to be clear.]] A total of 180 CCFs have been analysed and Fig. 18 shows the correlation between γ the BIS SPAN with a linear regression slope of 0.596. [[Referee: "Figure 9 shows the correlation between γ and BIS SPAN to understand the link Gram-

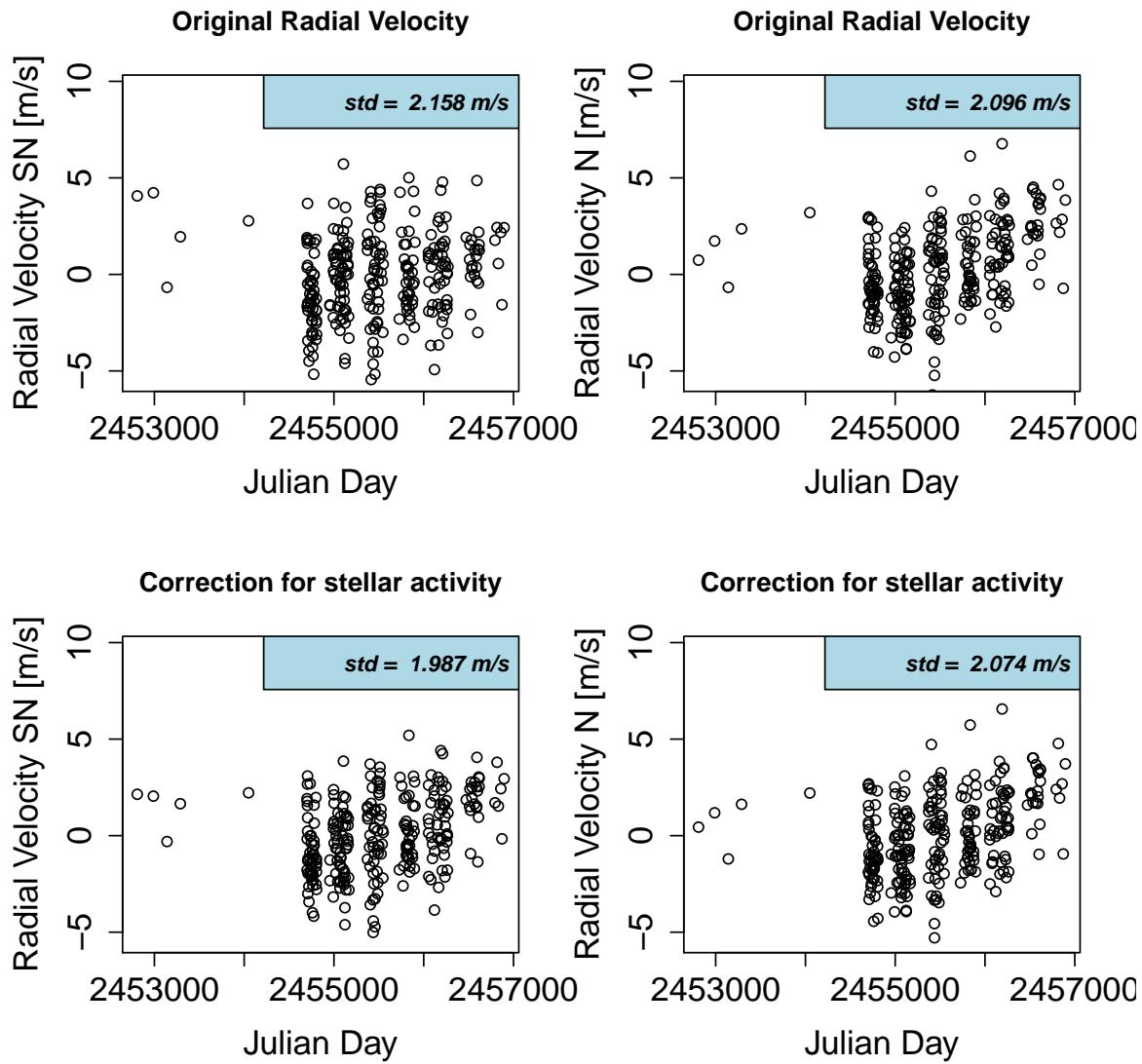


Figure 16: Radial Velocities for HD 215152 using the Normal and the Skew Normal fitting before and after correcting the analyses from the stellar activity.

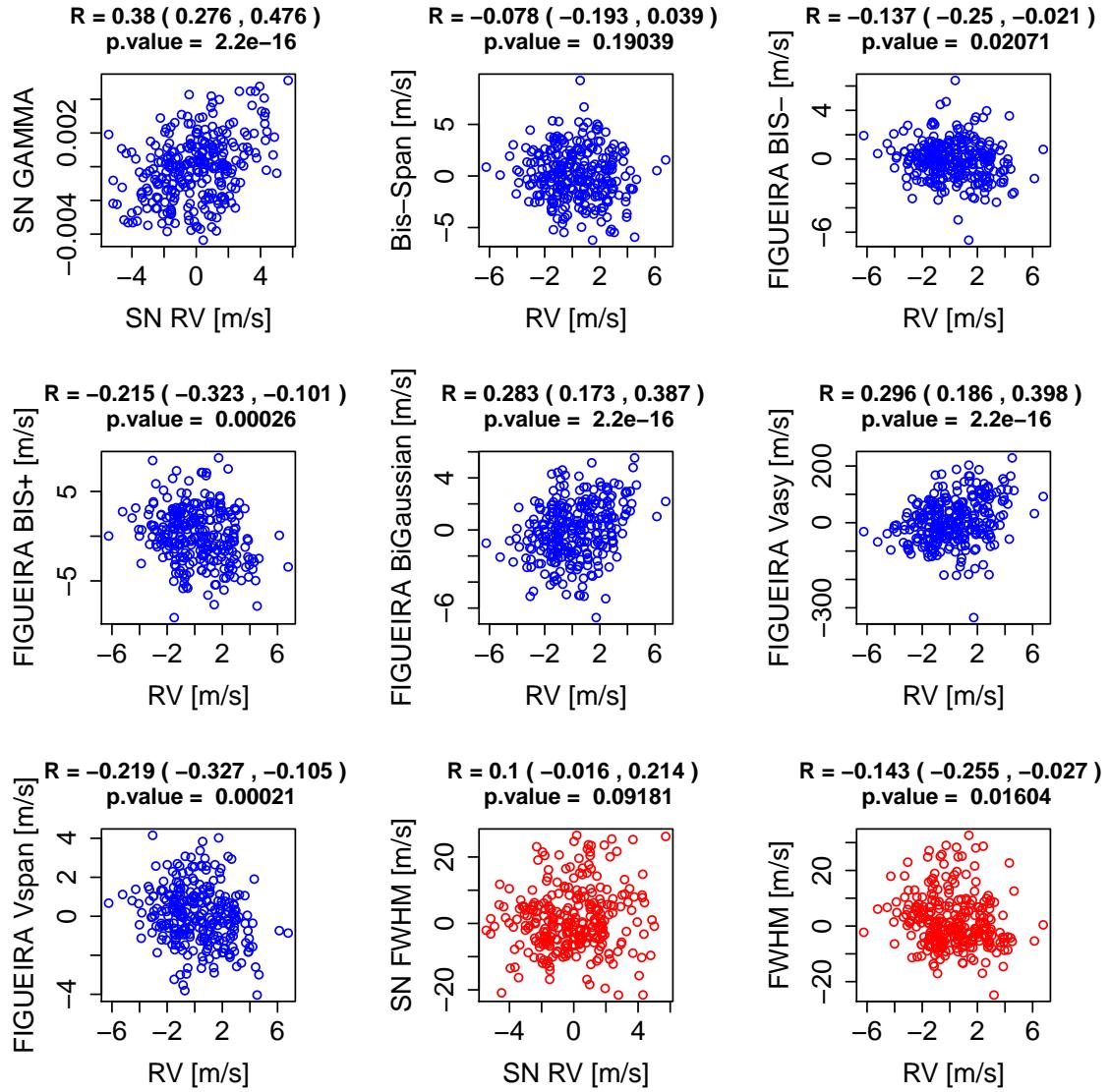


Figure 17: Correlation between the asymmetry parameters and the RVs for HD 215152. The last two plots show the correlation between the FWHM and the RVs for HD 215152 using respectively the SN and the Normal fits.

| Parameter | Normal Fitting | Skew Normal Fitting |
|-----------|----------------|---------------------|
| β_0 | 0.041 | 0.11 |
| β_1 | 0.79 | $4.12e - 11$ |
| β_2 | 0.042 | 0.11 |
| R^2 | 0.020 | 0.15 |

Table 4: **HD215152**: Evaluation of the linear combination used for correcting the RVs from stellar activity, according to Eq. 9. The p-values for the parameters β_0 , β_1 and β_2 for both the methodologies are summarized, as well as the R^2 . Concerning the Normal fitting, the intercept and the FWHM are statistically significant at explaining the RV variations at level 0.05 but not at level 0.01, explaining why the R^2 is only 0.02. The BIS SPAN is however not significant. For the SN case, it is the contrary, γ_1 is significant in explaining the SN RV variations, while the FWHM and the intercept are not. When looking at the value of R^2 , although the correlation is stronger for the SN case, is stay low at a level below 0.15.

| Parameter | Normal Fitting | Skew Normal Fitting |
|-----------|----------------|---------------------|
| β_0 | $2e - 16$ | $2.22e - 16$ |
| β_1 | 0.0067 | 0.00054 |
| β_2 | $2.22e - 16$ | $2.22e - 16$ |
| R^2 | 0.36 | 0.47 |

Table 5: **Corot-7**: Evaluation of the linear combination used for correcting the RVs from stellar activity, according to Eq. 9. The p-values for the parameters β_0 , β_1 and β_2 for both the methodologies are summarized, as well as the R^2 . For both the analyses, all the variables are helpful, although the linear combination is more useful in explaining the RV variability when working with the SN fitting.

matically incorrect sentence: "Fig. 9 shows to understand"? You need an alternate construction: "Figure 9 shows the correlation between BIS SPAN and ?? which helps us to understand...." Note: there is a correlation between two quantities which is not expressed in the original text.]]

The RVs obtained with the SN show more variability than the RVs derived with the Normal. [[Referee: Hasn't this been the case for virtually all the stars you looked at?]] After correcting for stellar activity using Eq. 9, the RV residuals with both distributions are similar. In Table 5, we see that in the Normal and the SN analyses, the intercept, the width and the asymmetry of the CCF (measured with FWHM or SN FWHM and γ or BIS SPAN) can explain part of the RV variability measured. Looking at R^2 , the RVs derived with the SN present like for all the stars in this paper a stronger correlation with γ_1 and SN FWHM than the RV derived in the Normal case.

The correlation between γ and the SN RVs of Corot-7 is much stronger than the correlation between the RVs and other asymmetric indicators (see Fig. 20). The correlation between γ and the SN RVs is 0.462, whereas the correlation between the other asymmetry statistics with their corresponding RVs result in values close to 0. Also the correlation between the FWHM and the

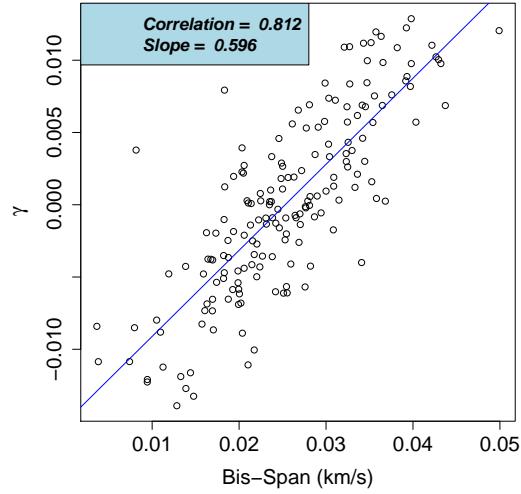


Figure 18: Correlation between γ and the BIS SPAN for Corot-7.

RVs of Corot-7 when using the SN fitting is stronger. These significant correlations for the SN parameters and non-significant for the Normal parameters suggest that for low signal to noise measurement, using the SN fitting improves our power in detecting stellar activity signals.

7 Estimation of standard errors for the CCF parameters

The HARPS reduction pipeline gives the pure photon-noise error estimate on the RVs. This error, called *noise*, is not derived from the Normal fit to the CCF but from the method described in ? using the CCF derivative. Then *noise* is associated to the RV error, $2 \times \text{noise}$ to the BIS SPAN error and $2.35 \times \text{noise}$ to the FWHM error.

In this section, we do a bootstrap analysis to measure the standard errors on the RVs or SN RVs, the FWHM or SN FWHM, and BIS SPAN or γ_1 . Because a CCF is obtained from a cross-correlation, each point in a CCF is correlated with each other. Therefore, we cannot do a bootstrap analysis on perturbing independently each CCF point with a Gaussian distribution scaled to the error of each given point. To bypass this problem, we bootstrap a hundred times the stellar spectrum given the photon-noise error of each wavelength and calculate for each realization a new CCF. We then fit a Normal or a SN to each of these CCFs, and the standard deviation of the distribution for the mean (RV or SN RV), the width parameter (FWHM or SN FWHM) and the asymmetric parameter (γ_1 or BIS SPAN) is associated to the error on each of these parameters. [[Referee: It is not clear what you are doing here. I can understand bootstrapping a time series: you shuffle all the values keeping the time stamps fixed. How do you bootstrap a stellar spectrum? Do you

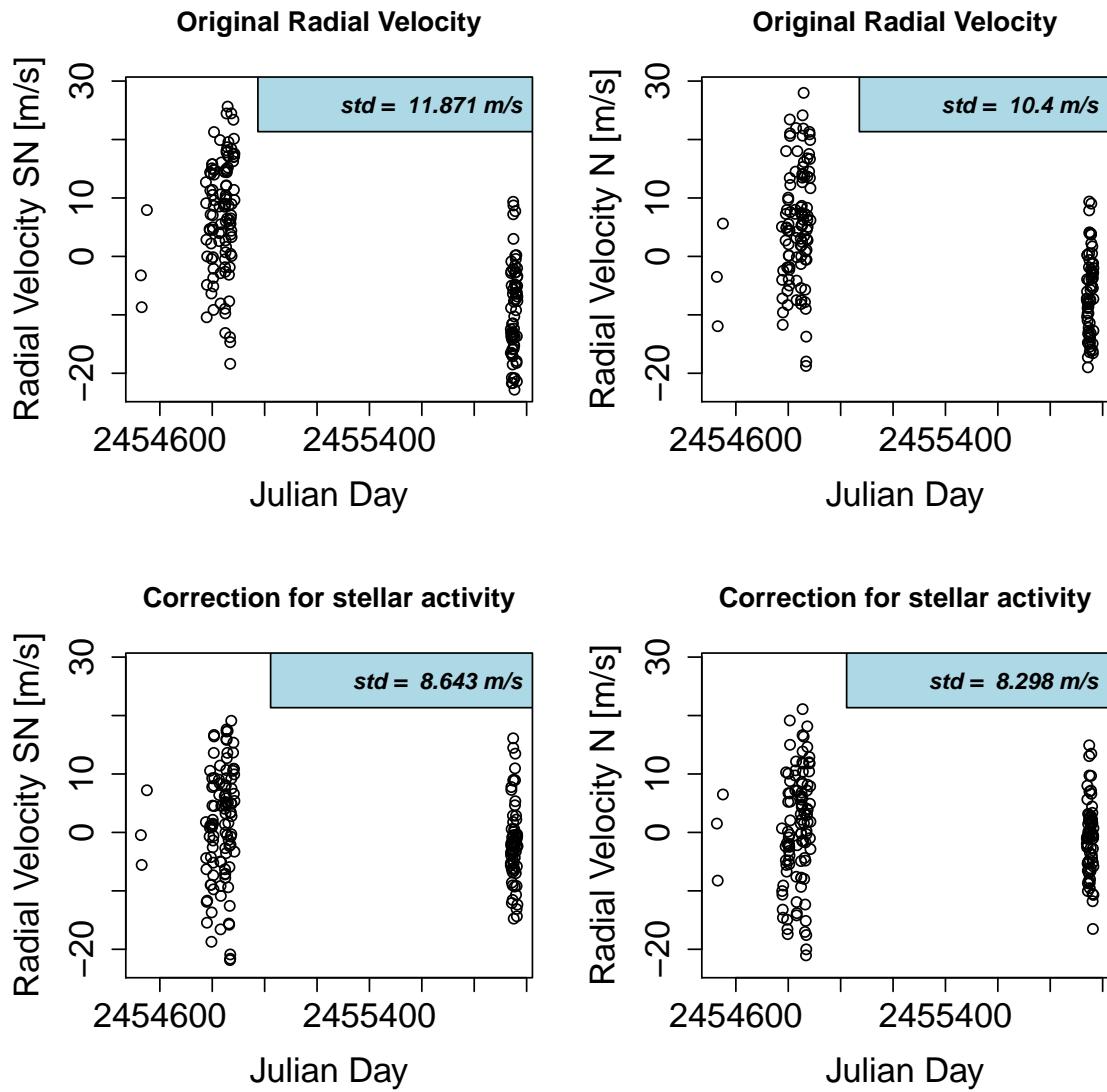


Figure 19: Radial velocities for Corot-7 using the Normal and the Skew Normal fitting before and after correcting the analyses from the stellar activity.

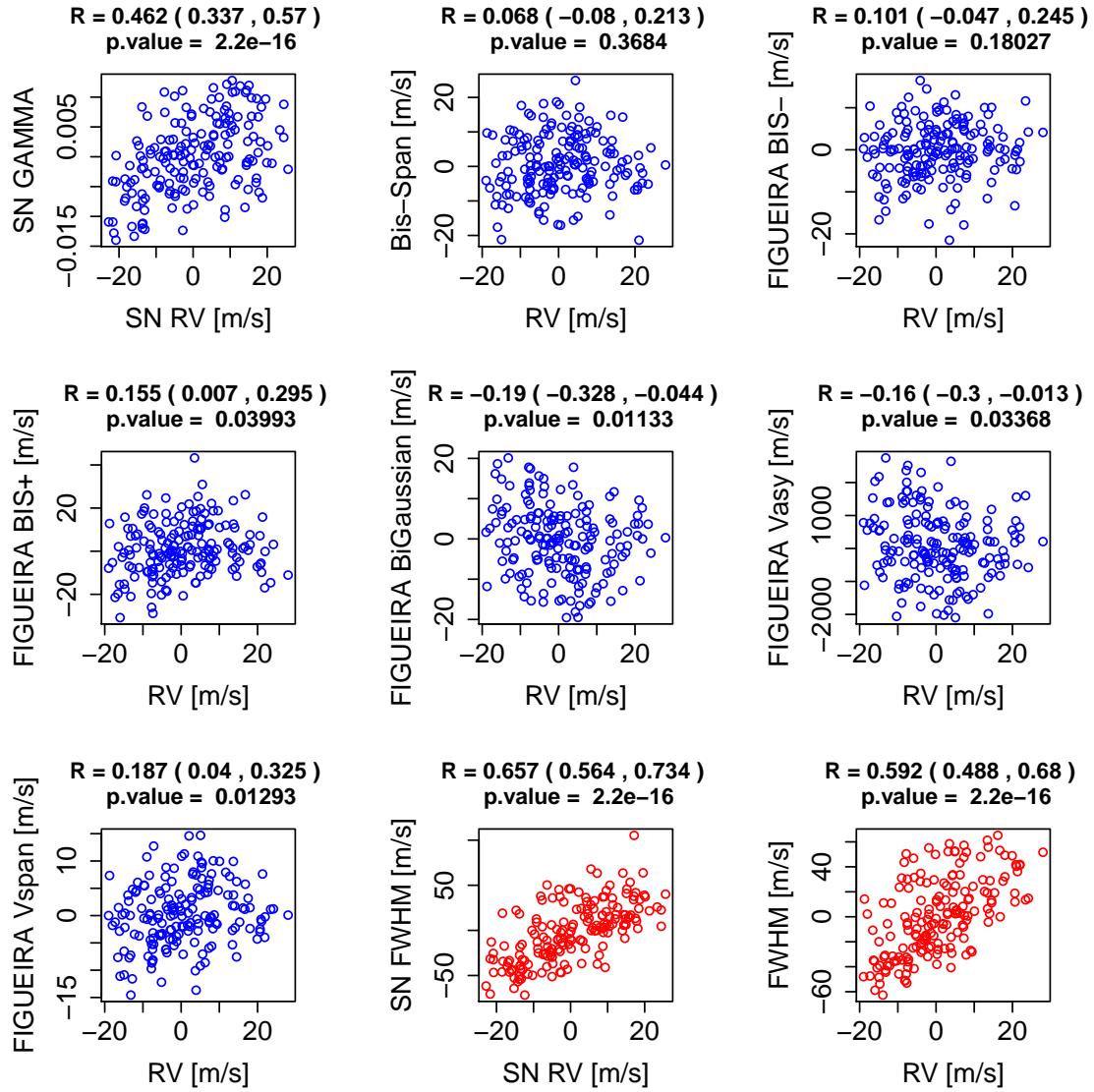


Figure 20: Correlation between the asymmetry parameters and the RVs for Corot-7. The last two plots show the correlation between the FWHM and the RVs for Corot-7 using respectively the SN and the Normal fits.

shuffle the intensity values keeping the wavelengths fixed? That would produce a mess. Do you just add different levels of random noise? But the spectra already have noise in them, they are real observations. You need to state more clearly here what you did.]]

In the top plots of Fig. 21 we show the different errors for the RVs, the width and the asymmetry of the CCFs for three star, HD215152, HD192310 and Corot-7, that are all at different SNR levels. The parameter SN50 corresponds to the SNR in order 50, which corresponds to a wavelength of 550 nm. In the bottom plots, we show the ratio between the parameters derived from the bootstrap analysis fitting the SN or the Normal and the *noise* parameter extracted from the HARPS reduction pipeline. We first see that the errors on the CCF parameters only depends on the SNR and do not depend on the spectral type. This is true if the spectral type are not too different though, like here where we show the results for G and K dwarfs.

We see that for the RV errors, the ratio between the *noise* parameter and the the RV error measured by the bootstrap using the Normal fitting is around 1, meaning that those two estimates of the RV errors are similar. However when fitting a SN to the CCF, we have RV errors that are on average 60% greater.

Thus, the RVs measured by the SN do not have the same precision than the RVs calculated using the Normal. This can be explained by the fact that in the case of the SN, variations in the asymmetry parameter γ_1 induce variations in the SN RV measured due to the strong correlation between those two parameters. Therefore, the errors in γ_1 propagate to the error in the SN RVs. [[Referee: This seems to be the case in all your stars. I do not consider a method that delivers errors that are 60% larger to be a very effective method.]] Regarding the errors in width of the CCF, we see that the bootstrap analysis for the Normal or the SN are equivalent to $2.35 \times \text{noise}$. Therefore, the precision in CCF width is the same if we fit a Normal or a SN to the CCF.

Finally, for the errors in CCF asymmetry, we see that the bootstrap analysis using the Normal fitting gives the same precision as using $2 \times \text{noise}$, which is the common errors used for BIS SPAN. However, when fitting the SN to the CCF, the asymmetry errors are 15% smaller. Therefore, the SN fit gives a better precision in CCF asymmetry than what can be reached using BIS SPAN.

Another interesting point is the behavior of the ratio of the bootstrap errors over the *noise* parameter as a function of SNR (see bottom plots in Fig. 21 for the RVs, CCF width and CCF asymmetry). We see that at low SNR, the bootstrap analysis gives a better precision, which can be explained by the fact that the *noise* parameter is estimated using the CCF derivative. Calculating the derivative of the CCF increases the noise and therefore explain this behavior at low SNR. For SNR below 200, the bootstrap analysis gives a better precision than the *noise* parameter.

8 Discussion

An analysis of the CCF residuals after fitting a Normal or SN distribution show that the SN is a better model to explain the shape of the CCF. This comes from the fact that CCFs present a natural asymmetry due the convective blueshift.

We compared for five stars the difference between the RVs, FWHM and asymmetry (BIS SPAN

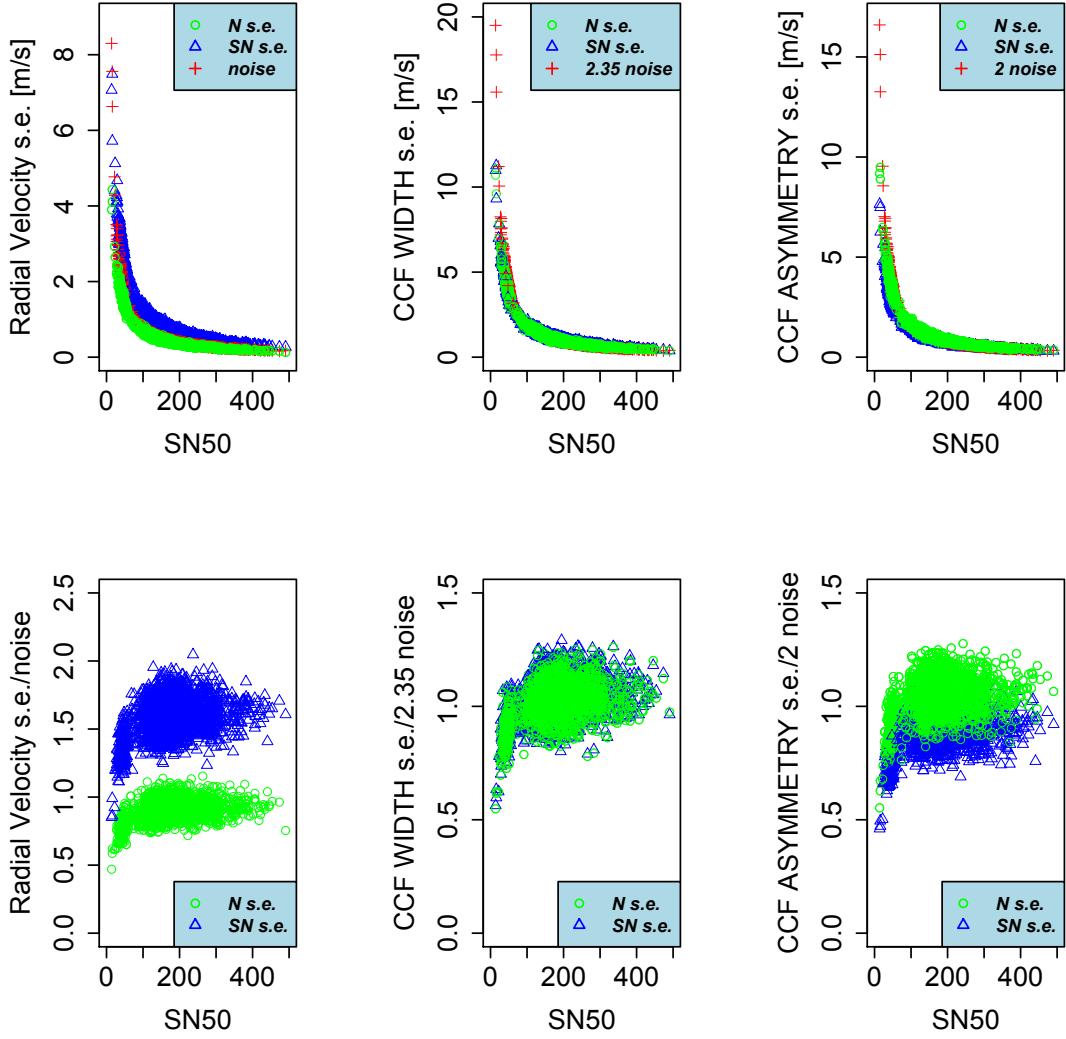


Figure 21: Comparison between the standard errors using the bootstrap analysis and the parameter *noise* for the RVs, $2.35 \times \text{noise}$ for the width, and $2 \times \text{noise}$ for the asymmetry parameter. Note that for the asymmetry, the error in BIS SPAN is in km s^{-1} . To be able to compare the errors in γ_1 and BIS SPAN, we multiplied the error in γ_1 by the slope of the correlation between γ_1 and BIS SPAN.

in the Normal case and parameter γ_1 in the SN case) as measured we a Normal or a SN fitting to the CCF. The γ_1 parameter is linearly dependant on the BIS SPAN, with always a strong correlation coefficient. The slope of this linear correlation change depending on the star studied. This is probably because the spectral type is different, therefore the effects from stellar activity are different.

When comparing the RV as measured by the Normal and the SN fitting, we always see that the rms of the latter is larger. Thus fitting the SN to the CCF gives RV measurements that are more sensitive to activity. This comes from the fact the the asymmetry parameter of the SN will modify as well the mean of the distribution, thus the RV. This is a positive point as it is easier to correct for a signal that is at a high SNR rather than a signal that is at a SNR of one, which is generally the case for the low-activity stars that are followed for RV surveys.

The RVs derived using the SN fitting present a higher sensitivity to stellar activity. However, once correcting the RVs from stellar activity using a linear combination of the width and asymmetry parameter of the CCF (FWHM and BIS SPAN in the Normal case, and SN FWHM and γ_1 in the SN case), the RV residuals in both the Normal and SN case show similar scatter. Therefore, the SN fitting does not improve stellar activity correction when using a simple linear combination with the width and asymmetry of the CCF. **[[Referee: In this statement you have basically admitted that your method is not very useful. I do not see why anyone would be compelled to use it! So, the bottom line is that you have introduced a method which produces similar results to the standard one. The only useful result of this paper is to convey to the readers that there is no significant advantage in implementing the SN method!]]** However, because the amplitude of the variations is larger in the raw SN RVs, correction techniques based on Gaussian Processes might be more efficient (??).

When looking at the correlation between the asymmetry and width parameters of the CCF (FWHM and BIS SPAN or the alternative indicators in ? in the Normal case, and SN FWHM and γ_1 in the SN case) with respect to the RVs (RVs in the Normal case or SN RVs in the SN case), we observe that the correlations are always stronger for the parameters of the SN. Therefore, the SN parameters are more sensitive to activity. In the case of Tau Ceti, which is at very low activity level, we find a significant correlation of 0.57 between γ_1 and SN RV, while for all the other asymmetric parameterization, BIS SPAN or the alternative indicators in ?, the correlations are weaker with a maximum of 0.21.

We also studied the behavior of the noise on the different parameters of the Normal and the SN as a function of SNR. By performing a bootstrap analysis, we confirmed that in the case of the Normal, using the *noise* parameter as returned by the HARPS pipeline gives a correct estimate of the errors measured on the RVs, FWHM and BIS SPAN for $\text{SNR} > 150$. For the RVs, the errors are equivalent to *noise*, for the FWHM to $2.35 \times \text{noise}$ and for the BIS SPAN to $2 \times \text{noise}$. For lower SNR though, the bootstrap analysis gives better results, and this comes from the fact the the *noise* parameter is estimated using the derivative of the CCF, which increase the noise (?). For very low SNR, the improvement of the bootstrap analysis can be as high as 50%. Therefore, although taking much more time, we encourage the use of bootstrapping when calculating errors for the different parameters of the CCF.

In the case of the errors when using the SN fitting, the precision on the SN RVs is however

60% worse than for the RVs. This comes from the fact that because the asymmetry is accounted for in the SN and because the SN RVs and γ_1 are strongly correlated, the errors in γ_1 propagates to the errors in SN RVs. This is not so important for $\text{SNR} > 150$ for which the errors are below the m s^{-1} precision of HARPS and the m s^{-1} variation induced by stellar activity signals even in low activity cases². These larger errors in RVs can be a problem for low SNR measurements. In the case of the CCF width, the error derived when using the Normal or SN fitting are very similar. Finally for the asymmetry, we see that the error in γ_1 is $\sim 15\%$ smaller than the error in BIS SPAN.

9 Conclusion

In this paper we introduced a novel approach based on the Skew Normal (SN) distribution for deriving RVs and shape variations in the CCF of stars. When searching for small-mass exoplanets using the RV technique, it is essential to understand the shape variation of the CCF, which is a proxy for stellar activity effects. The standard approach consist in first to adjust a Normal distribution to the CCF to get the RV and FWHM, defined as the mean and the FWHM of the Normal distribution, and then to measure the asymmetry by calculating BIS SPAN. FWHM and BIS SPAN give us information on the line shape that are used to probe stellar activity signals. In this paper we demonstrate that by using the SN distribution to fit CCFs, we can measure simultaneously the RV of the star with the width and asymmetry of the CCF.

Using the SN to fit CCFs brings a significant improvement in probing stellar activity. The correlations between SN RV and SN FWHM, and SN RV and the asymmetric parameter γ_1 are much stronger than the correlations between the equivalent parameters derived using a Normal fit (RV, FWHM and BIS SPAN or the asymmetric parameters described in ?). The RVs derived by the SN are also more sensitive to activity and the precision on the asymmetry measured by γ_1 is greater than the one on BIS SPAN by $\sim 15\%$. Therefore when searching for rotational periods in the data, or applying Gaussian Processes to account for stellar activity signals, the SN parameters should be used.

At first glance, one downside of using the SN comes from the RV errors that are greater by 60% relative to the RV errors measured when considering a Normal. This is due to the strong correlation between SN RV and γ_1 and this can be a problem for low SNR measurements. However, in the regime of $\text{SNR} > 150$, which is the aim of all RV surveys at high precision, the precision of the SN RVs is below the m s^{-1} , which is thus below the instrumental precision of HARPS and below the RV perturbations induced by stellar activity.

Finally, we also encourage the use of bootstrapping to estimate more realistic errors on the different parameters of the Normal or SN fitted to the CCF, mainly in the low SNR regime where a gain of 50% can be reached. This takes significantly more time, but note that 100 realization are enough to get a good estimation of errors.

²As we can see in papers publishing HARPS results using an additional jitter to consider stellar activity noise not accounted for in the *noise* parameter, the extra jitter is always greater than 0.8 m s^{-1} (e.g. ?).

10 Acknowledgements

We are grateful to all technical and scientific collaborators of the HARPS Consortium, ESO Headquarters and ESO La Silla who have contributed with their extraordinary passion and valuable work to the success of the HARPS project. XD is grateful to the Society in Science–The Branco Weiss Fellowship for its financial support.